ERDC/CHL CHETN-X-1
February 2016

US Army Corps of Engineers®

A Bayesian Analysis of the Flood Frequency Hydrology Concept

by Brian E. Skahill, Alberto Viglione, and Aaron Byrd

PURPOSE: The purpose of this document is to demonstrate a Bayesian analysis of the flood frequency hydrology concept as a formal probabilistic-based means by which to coherently combine and also evaluate the worth of different types of additional data (i.e., temporal, spatial, and causal) in a flood frequency analysis. This approach is responsive to the stated ultimate goal of existing U.S. Army Corps of Engineers (USACE) policy guidance, which is probabilistic analysis of "all key variables, parameters, and components of flood damage reduction studies" (USACE 2006). This objective will be accomplished by independently revisiting components of an example originally profiled by Viglione et al. (2013). This technical note will also include a brief discussion of some potential opportunities for future related research and development.

INTRODUCTION: Merz and Blöschl (2008a,b) proposed the concept of flood frequency hydrology, which emphasizes the importance of combining local flood data with additional types of temporal, spatial, and causal information using hydrologic reasoning to perform a flood frequency analysis at a site of interest. Temporal expansion involves the collection and consideration of information on flood behavior before or after the period of record of measured discharge. It accommodates short records that are not completely representative of a system's flood behavior. Flood marks on buildings and paleoflood information are two types of temporal information expansion data. Spatial information expansion involves trading space for time by using flood information from neighboring systems, viz., a regional flood frequency analysis methodology such as the index flood method (Dalrymple 1960) to improve upon the flood frequency analysis at the site of interest. Introducing hydrologic understanding of local flood production factors is the goal of causal information expansion. The derived flood frequency approach (e.g., Eagleson 1972; Sivapalan et al. 1990; Rahman et al. 2002; Sivapalan et al. 2005), the Gradex method (Guillot 1972; Duband et al. 1994; Naghettini et al. 1996), and rainfall-runoff modeling are all examples of causal information expansion.

With flood frequency hydrology, in estimating flood frequencies, the intent is to extract the maximum amount of information from all available complementary data sources and to combine the additional data types (i.e., temporal, spatial, and causal) using hydrologic reasoning. Merz and Blöschl (2008a,b) underscore that a key element of the combination process is to account for the uncertainty of the various pieces of information. Whereas Merz and Blöschl (2008a,b) relied upon heuristic hydrologic reasoning to combine the different data types, Viglione et al. (2013) revisited the flood frequency hydrology concept within a Bayesian analysis framework. In particular, they profiled the flood frequency hydrology concept by employing a Metropolis-Hastings (Metropolis et al. 1953; Hastings 1970) jumping rule based Bayesian Markov Chain Monte Carlo (MCMC) sampler to simultaneously optimize and infer the generalized extreme value (GEV) distribution parameters using a systematic discharge record, a systematic record plus one of each form of

information expansion (i.e., temporal, spatial, or causal), and a systematic record plus all forms of information expansion.

Markov Chain Monte Carlo (MCMC) simulation is a formal Bayesian approach for estimating the posterior probability distribution of the specified adjustable model parameters, in this case, the GEV distribution parameters. It treats the specified adjustable model parameters as random variables and relies upon Bayes' Theorem to compute their joint posterior probability distribution. Bayes' Theorem effectively communicates that the posterior distribution is proportional to the product of the prior distribution, prescribed based on the modeler's best judgment, expert opinion, or literature estimates, among possible others, and the likelihood function (i.e., the conditional distribution), which encapsulates the conditioning process with the observed dataset, which in this case is a systematic record of annual discharge maxima plus possibly one or more forms of information expansion. The idea behind MCMC simulation is that while one wants to compute a probability density, $p(\mathbf{p}|D)$, where $\mathbf{p}$ and D represent the vector of adjustable model parameters and the data/information imparted to the analysis, respectively, there is the understanding that such an endeavor may be impracticable. Additionally, simply being able to generate a large random sample from the probability density would be equally sufficient as knowing its exact form. Hence, the problem then becomes one of effectively and efficiently generating a large number of random draws from $p(\mathbf{p}|D)$. It was discovered that an efficient means to this end is to construct a Markov chain, a stochastic process of values that unfold in time, with the following properties: (1) the state space (set of possible values) for the Markov chain is the same as that for $\mathbf{p}$; (2) the Markov chain is easy to simulate from; and (3) the Markov chain's equilibrium distribution is the desired probability density $p(\mathbf{p}|D)$. The Gelman and Rubin (1992) quantitative measure is commonly employed to assist with diagnosis of chain convergence. A Markov chain with the above-mentioned properties can be constructed by choosing a symmetric proposal distribution and employing the Metropolis acceptance probability (Metropolis et al. 1953) to accept or reject candidate points. By constructing such a Markov chain, one can then run it to equilibrium (and this period is often referred to as the sampler *burn-in* period) and subsequently sample from its stationary distribution. Within the context of its application to simultaneously optimize and infer the GEV distribution parameters using a systematic record and one or more forms of information expansion, the post burn-in random draws from p can be used to construct credible intervals for the estimated flood quantiles.

Hence, by performing the flood frequency hydrology concept within a Bayesian analysis framework, a formal probabilistic-based and flexible means is employed not only for simultaneous optimization and inference but also for combining the different data types via application of Bayes' theorem. A Bayesian analysis of the flood frequency hydrology concept dovetails with the stated goal of existing related USACE policy guidance. In particular, the USACE is required to perform risk and uncertainty analyses in the process of planning, design, and operation of all civil works flood risk management projects as described in Engineer Regulation (ER) 1105-2-101 (USACE 2006) and its cited references (e.g., Engineer Manual [EM] 1110-2-1619 [USACE 1996]). The risk-informed analysis framework presented in ER 1105-2-101 (USACE 2006), jointly promulgated by the USACE Planning and Engineering communities of practice, requires acknowledgement of and accounting for error and uncertainty in the "key variables, factors, parameters, and data components" relevant to the planning and design of flood damage reduction projects. By capturing and quantifying "the extent of the risk

and uncertainty in the various planning and design components of an investment project," it permits for an evaluation of the tradeoff between risks and costs.

The Bayesian analysis of the flood frequency hydrology concept performed by Viglione et al. (2013) is independently revisited in this technical note using an adaptive population-based MCMC sampler (ter Braak and Vrugt 2008). The revisited Bayesian analysis contained in this technical note demonstrates a USACE capacity to combine additional information beyond that of the systematic record into a flood frequency analysis. The additional data types considered herein include historical and causal forms of information expansion. The casual information expansion data were derived by way of expert elicitation and in a formal probability-based fashion rather than simply arbitrarily. The systematic record and the additional data types imparted to the analysis are flexibly combined in an easily revisable manner that is consistent, throughout the entire analysis framework, with the previously mentioned need for probabilistic analysis for flood damage reduction studies within the USACE.

The remainder of this technical note independently revisits pieces of the Bayesian analysis of the flood frequency hydrology concept performed by Viglione et al. (2013) for the 622 km^2 Kamp at Zwettl river basin located in northern Austria. It not only underscores attributes of the method as applied to the Kamp at Zwettl but also discusses ways in which the approach compares with current practice. Moreover, it concludes by expressing some opportunities for related research and development.

EXAMPLE: The 55-year record (1951–2005) of available annual discharge maxima for the Kamp at Zwettl river basin is of great interest by virtue of the 2002 extreme flood event. Excluding the 2002 flood by only considering the first 51 years of the systematic record results in an estimate for the 100-year flood runoff (Q_{100}) of 159 m^3/s and an assigned return period for the 2002 flood greater than 100,000 years. Whereas, the estimate for Q_{100} is 285 m^3/s, and the 2002 flood is assigned a return period of 340 years when all 55 years are employed to fit the GEV distribution parameters using the method of L-moments. Viglione et al. (2013) explored the flood frequency hydrology concept, via a Bayesian analysis, not only considering the first 51 years but also the entire 55 years of the available systematic record to examine how well a flood of the magnitude of the 2002 event could be anticipated statistically prior to its occurrence. Elements of that complete analysis are revisited herein via application of Bayesian MCMC, not only considering the systematic record before and after the 2002 flood but also temporal information expansion, causal information expansion, and a combination of the temporal and causal information expansions. In particular, eight distinct primary MCMC simulations were performed, as listed in Table 1, to simultaneously optimize and infer the GEV distribution parameters using data of the Kamp at Zwettl. Viglione et al. (2013) summarize the assumptions made in the Bayesian analysis.

Systematic Data. The first two MCMC simulations listed in Table 1 solely consider the systematic record for the Kamp at Zwettl, either up to 2001, just before the 2002 flood event (i.e., 1951–2001), or the complete available record (i.e., 1951–2005). In either case, an uninformed uniform prior distribution is employed as well as a likelihood function of the form

$$l\left(D\middle|\mathbf{p}\right) = l_s\left(D\middle|\mathbf{p}\right) = \prod_{i=1}^{s} f_x\left(x_i\middle|\mathbf{p}\right) \tag{1}$$

Table 1. Summary of Kamp at Zwettl data employed for each of the eight distinct MCMC simulations.

MCMC simulation	Data of the Kamp at Zwettl
1	Systematic data (1951-2001)
2	Systematic data (1951-2005)
3	Systematic data (1951-2001) + temporal information expansion
4	Systematic data (1951-2005) + temporal information expansion
5	Systematic data (1951-2001) + causal information expansion
6	Systematic data (1951-2005) + causal information expansion
7	Systematic data (1951-2001) + temporal + causal information expansion
8	Systematic data (1951-2005) + temporal + causal information expansion

where D is the sample set of recorded annual discharge maxima, x_i; s is the record size; and f is the three parameter $\mathbf{p} = (p_1, p_2, p_3)$ GEV distribution (p_1, p_2, and p_3 represent the location, shape, and scale parameters, respectively). The flood frequency curves presented in Figure 1 and Figure 2 were obtained by applying MCMC using the systematic data up to 2001, and also the entire systematic data record, respectively. In each figure, the flood frequency curve estimates shown correspond to the posterior mode (PM) (i.e., the GEV with $\mathbf{p}$ that maximizes $p(\mathbf{p}|D)$) and the computed 90% credible intervals, which are subdomains of the predictive distributions, characterized by the postburn-in random draws, for a given return period or peak discharge. For each simulation, Table 2 lists the PM estimates for the GEV parameters and also for Q_{100}, Q_{1000}, and their corresponding computed 90% credible intervals. The probability that Q_{100}/Q_{1000} lies within the 90% credible bounds specified in Table 2, given the data to support each distinct Bayesian analysis, is in each case 0.9.

Temporal Information Expansion. The next two MCMC simulations listed in Table 1 (i.e., simulations 3 and 4) that involve temporal information expansion considered three historical floods (y_1, y_2, and y_3; $k = 3$) that occurred in 1655, 1803, and 1829 during the historical period of 1600 through 1950. It is assumed that the specified threshold of $X_0 = 300$ m^3/s is only exceeded k times during the defined historical period of h years (h = 350 = 1950 – 1600). The specified perception threshold is the maximum possible value of the smallest of the three historic events. Uncertainty bounds given by $\left[y_{L_j}, y_{U_j}\right]$ for $j = 1,...,k$, and equal to $\pm 25\%$ of the estimated peak discharges, based on expert judgment, are designated for each historic event (Wiesbauer 2004, 2007). The likelihood function representing the joint probability of occurrence of recent (l_s) and historical (l_H) flood observations is given by

$$l\left(D\middle|p\right) = l_s\left(D\middle|p\right) \bullet l_H\left(D\middle|p\right) \tag{2}$$

where l_s is given by equation (1), F below denotes the cumulative of f, and

$$l_H\left(D\middle|\mathbf{p}\right) = \binom{h}{k} F_x\left(X_0\middle|\mathbf{p}\right)^{(h-k)} \left\{\prod_{j=1}^{k}\left[F_x\left(y_{u_j}\middle|\mathbf{p}\right) - F_x\left(y_{L_j}\middle|\mathbf{p}\right)\right]\right\} \tag{3}$$

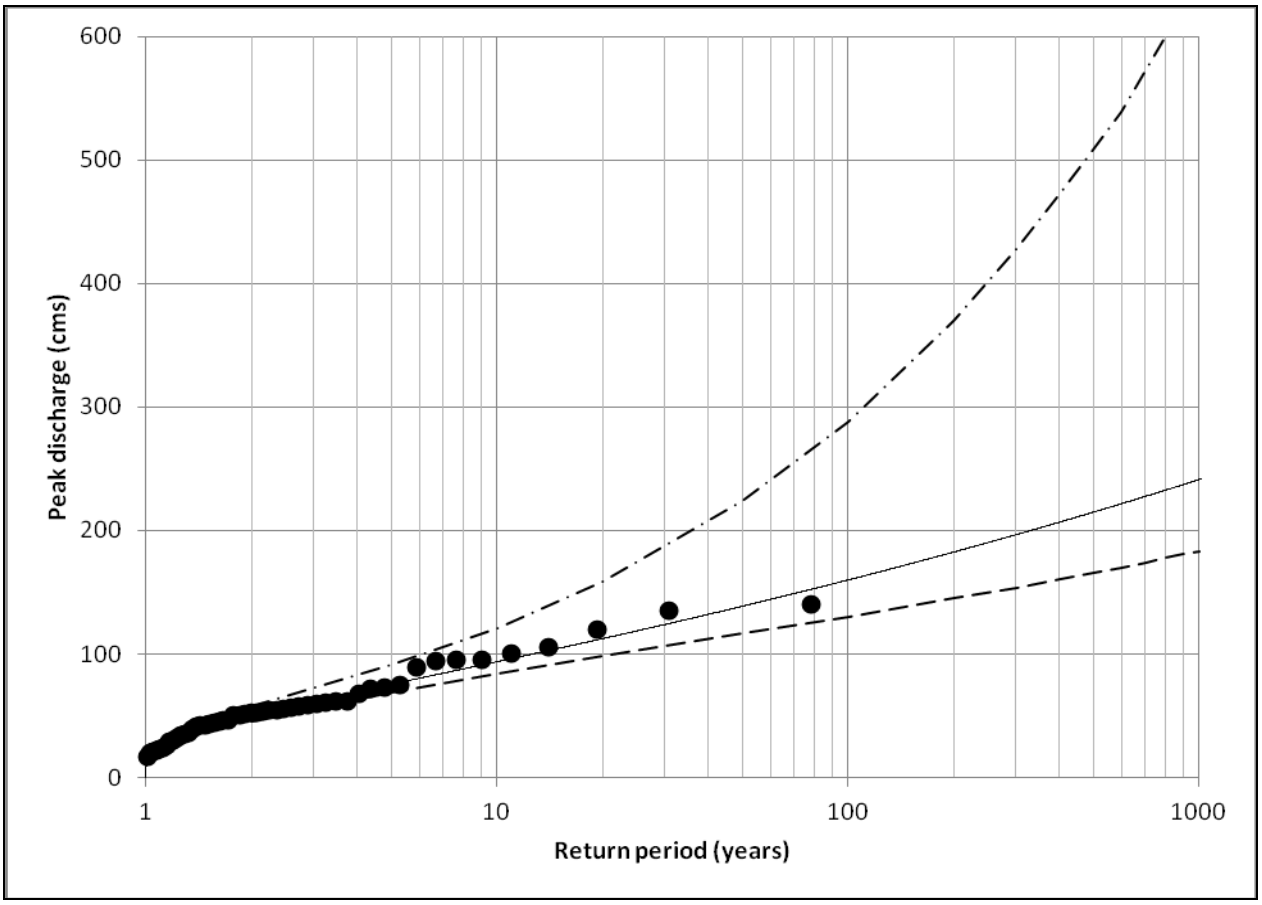

Figure 1. Bayesian fit of the GEV distribution to the systematic data record of the Kamp at Zwettl for 1951–2001. The continuous line corresponds to the PM estimate. The 5% and 95% credible bounds are shown as dashed lines. (cms = m^3/s)

Table 2. For each MCMC simulation, the computed PM estimate for the GEV parameters and also for Q100 and Q1000, including their 90% credible bounds, at the Kamp. The last column includes the PM-based estimate for the return period, T, in years for a flood equal in magnitude to the 2002 flood event.

MCMC simulation	GEV Parameters			Q100 (m3/s)			Q1000 (m3/s)			T (years)
	p1	p2	p3	PM	5%	95%	PM	5%	95%	PM
1	42.9	20.2	-0.096	160	130	288	241	183	649	84674
2	41.7	20.7	-0.310	253	184	542	543	317	1853	598
3	43.4	21.7	-0.222	217	176	291	399	278	647	1752
4	42.6	21.5	-0.281	244	197	331	497	347	818	767
5	41.9	21.0	-0.313	258	193	307	557	335	702	557
6	41.6	20.8	-0.333	269	217	317	604	418	747	454
7	42.7	21.8	-0.291	253	206	299	527	369	671	643
8	42.5	21.5	-0.313	264	220	308	571	418	708	517
9	43.3	21.7	-0.234	223	179	293	419	287	653	1423
10	42.6	21.5	-0.288	249	201	323	514	352	775	694
11	42.5	21.7	-0.322	271	251	298	598	532	647	458
12	42.4	21.6	-0.325	272	252	298	602	538	653	451

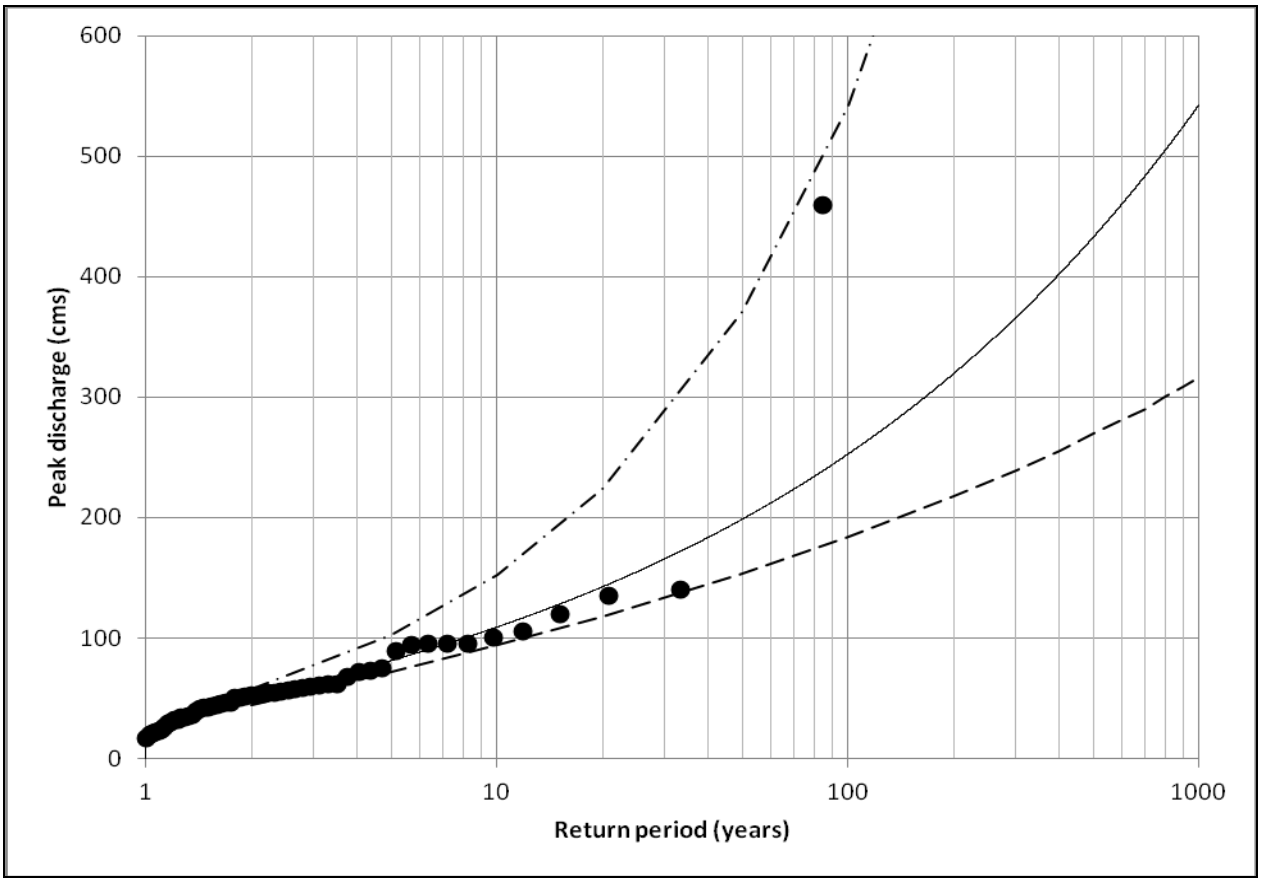

Figure 2. Bayesian fit of the GEV distribution to the systematic data record of the Kamp at Zwettl for 1951–2005. The continuous line corresponds to the PM estimate. The 5% and 95% credible bounds are shown as dashed lines. (cms = m^3/s)

The likelihood function $l\left(D\middle|\mathbf{p}\right)$ of Equation (2) combines three terms: (1) the probability density function of the s systematic data; (2) the probability of observing no events above the perception threshold for $h-k$ years; and (3) the probability of observing k historical events between the specified lower and upper bounds. As with the first two MCMC simulations, an uninformative uniform prior is utilized. The flood frequency curves presented in Figure 3 and Figure 4 were obtained via application of MCMC and considering the systematic and temporal information expansion data until 2001 and until 2005, respectively. Table 2 lists the PM estimates for the GEV parameters and also for Q_{100}, Q_{1000}, and their corresponding computed 90% credible intervals.

Causal Information Expansion. Viglione et al. (2013) explored the inclusion of causal information expansion data within the flood frequency hydrology concept for the Kamp, via Bayesian analysis, by incorporating information derived from expert judgment regarding the 500-year flood peak. Coles and Tawn (1996) studied the elicitation and formulation of prior information for a Bayesian analysis of extreme rainfall. They elicited prior information in terms of extreme quantiles, arguing it to be far more realistic to expect an expert to meaningfully quantify their prior beliefs about extremal behavior rather than the distribution's parameters. For this analysis, expert elicitation, based on rainfall-runoff modeling with artificial rainfall series, and the

expert's familiarity with the system resulted in an estimate for Q_{500} of 480 m^3/s $\pm 20\%$, which was reformulated, working together with the expert, to be given by

$$k(Q_{500}) = N(\mu_{500}, \sigma_{500}) \tag{4}$$

where $\mu_{500} = 480$ m^3/s; $\sigma_{500} = 80$ m^3/s; and N denotes the normal distribution. The GEV quantile with 500-year return period is given by

$$Q_{500} = g(\mathbf{p}) = p_1 + \frac{p_2}{p_3}\left[1 - \left(ln\left(\frac{500}{500-1}\right)\right)^{p_3}\right] \tag{5}$$

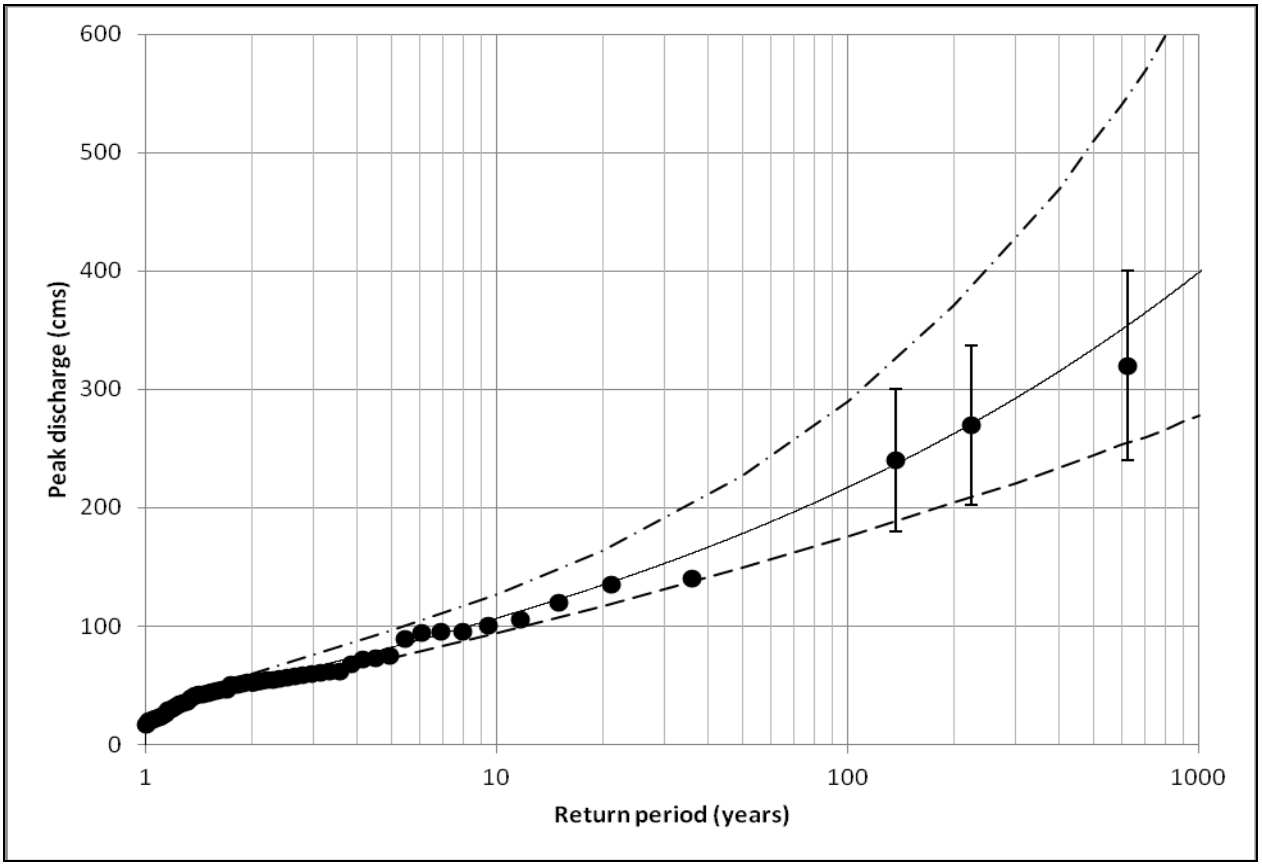

Figure 3. Bayesian fit of the GEV distribution to the systematic data record of the Kamp at Zwettl for 1951–2001, including temporal information expansion data, also shown. The continuous line corresponds to the PM estimate. The 5% and 95% credible bounds are shown as dashed lines. (cms = m^3/s)

In this case, during MCMC simulation, the likelihood function $l(D|\mathbf{p})$ is multiplied by $k(g(\mathbf{p}))$ to calculate p. The flood frequency curves presented in Figures 5 and 6 were obtained via application of MCMC and considering the systematic and causal information expansion data until 2001 and until 2005, respectively. Table 2 lists the PM estimates for the GEV parameters and also for Q_{100}, Q_{1000}, and their corresponding computed 90% credible intervals.

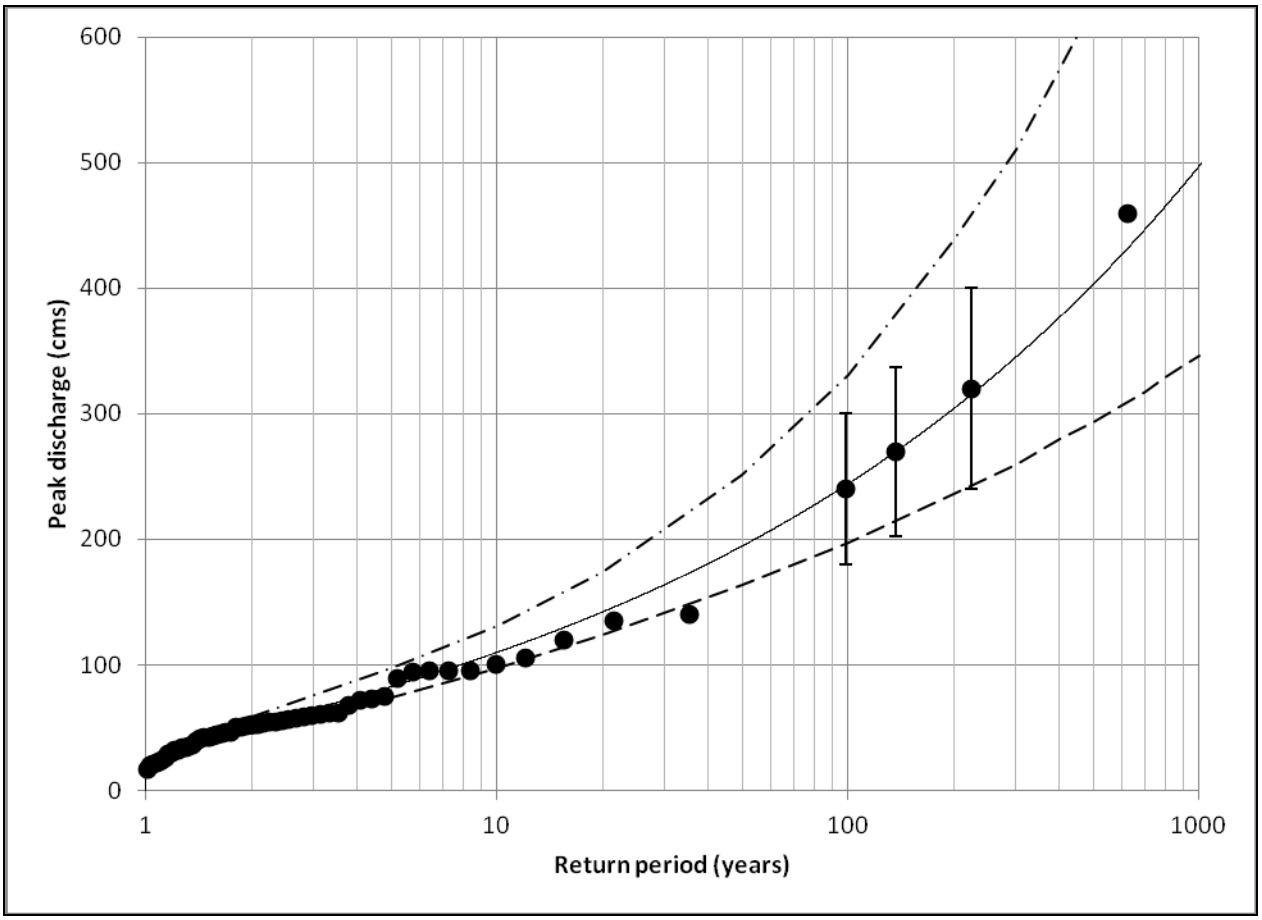

Figure 4. Bayesian fit of the GEV distribution to the systematic data record of the Kamp at Zwettl for 1951–2005, including temporal information expansion data, also shown. The continuous line corresponds to the PM estimate. The 5% and 95% credible bounds are shown as dashed lines. (cms = m^3/s)

Combination of Data. Viglione et al. (2013) underscore the appeal of a Bayesian analysis of the flood frequency hydrology concept, viz., the capacity to combine and account for all of the different information together. Bayesian MCMC was applied to combine the systematic data together with the temporal and causal information expansion data. In this case,

$$p\left(\mathbf{p}\middle|D\right) \propto l_s\left(D\middle|\mathbf{p}\right) \bullet l_H\left(D\middle|\mathbf{p}\right) \bullet k\left(g\left(\mathbf{p}\right)\right) \tag{6}$$

The results of these two MCMC simulations (i.e., 7 and 8) are shown in Figures 7 and 8, and also in Table 2. Four additional MCMC simulations were also performed to explore the impact of varying the standard deviation associated with the expert's estimate for Q_{500}. MCMC simulations 9/11 and 10/12 combined the systematic data record (until 2001 and also until 2005) together with the temporal and causal information expansion data considering a value of $\sigma_{500} = 240/26.7$ m^3/s for the causal information expansion data. The results of these four additional MCMC simulations are presented in Figures 9–12 and also in Table 2.

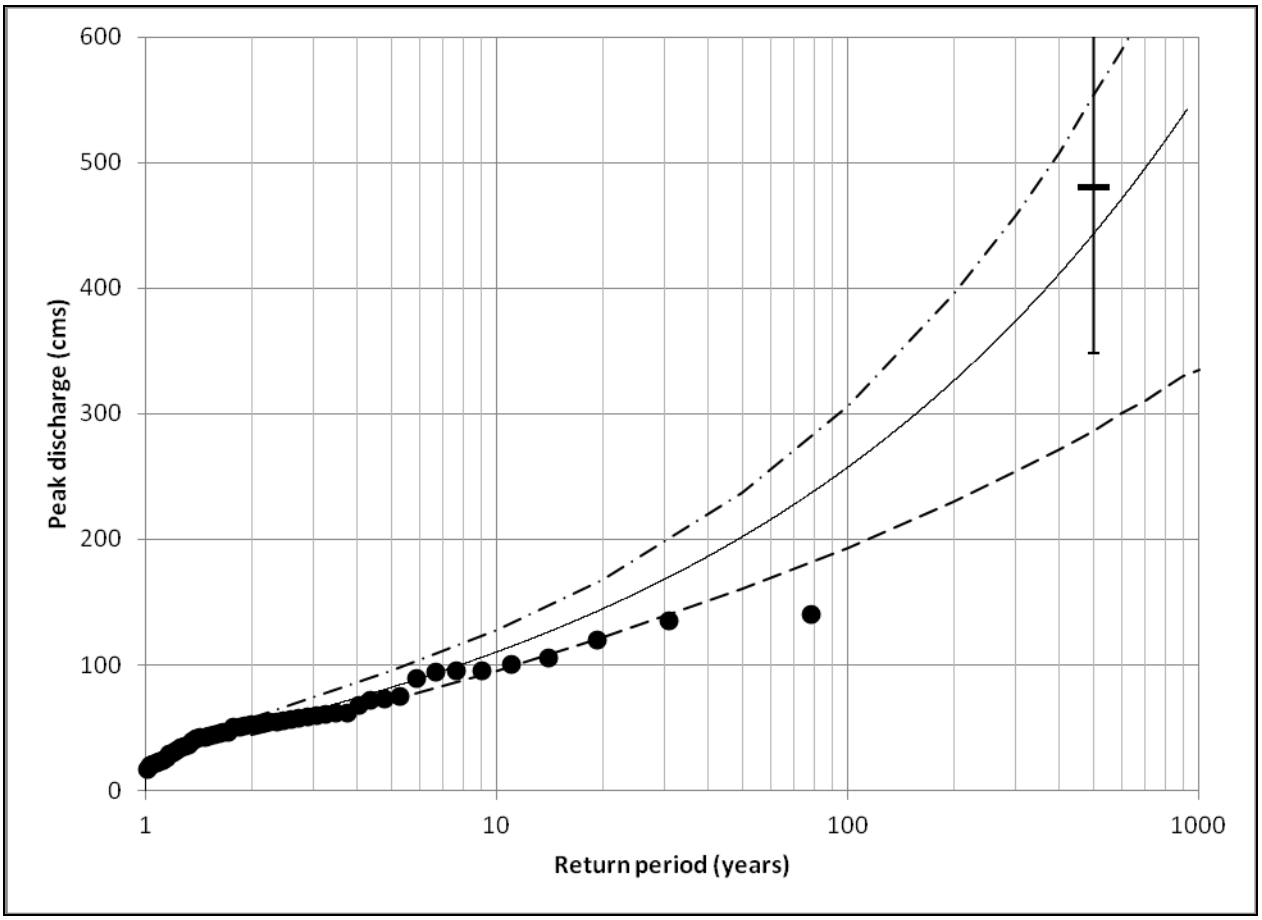

Figure 5. Bayesian fit of the GEV distribution to the systematic data record of the Kamp at Zwettl for 1951–2001, including causal information expansion. The expert guess for Q_{500} is shown along with its 5% and 95% quantiles ($\sigma_{500} = 80\,\mathrm{m^3/s}$). The continuous line corresponds to the PM estimate. The 5% and 95% credible bounds are shown as dashed lines. (cms = m^3/s)

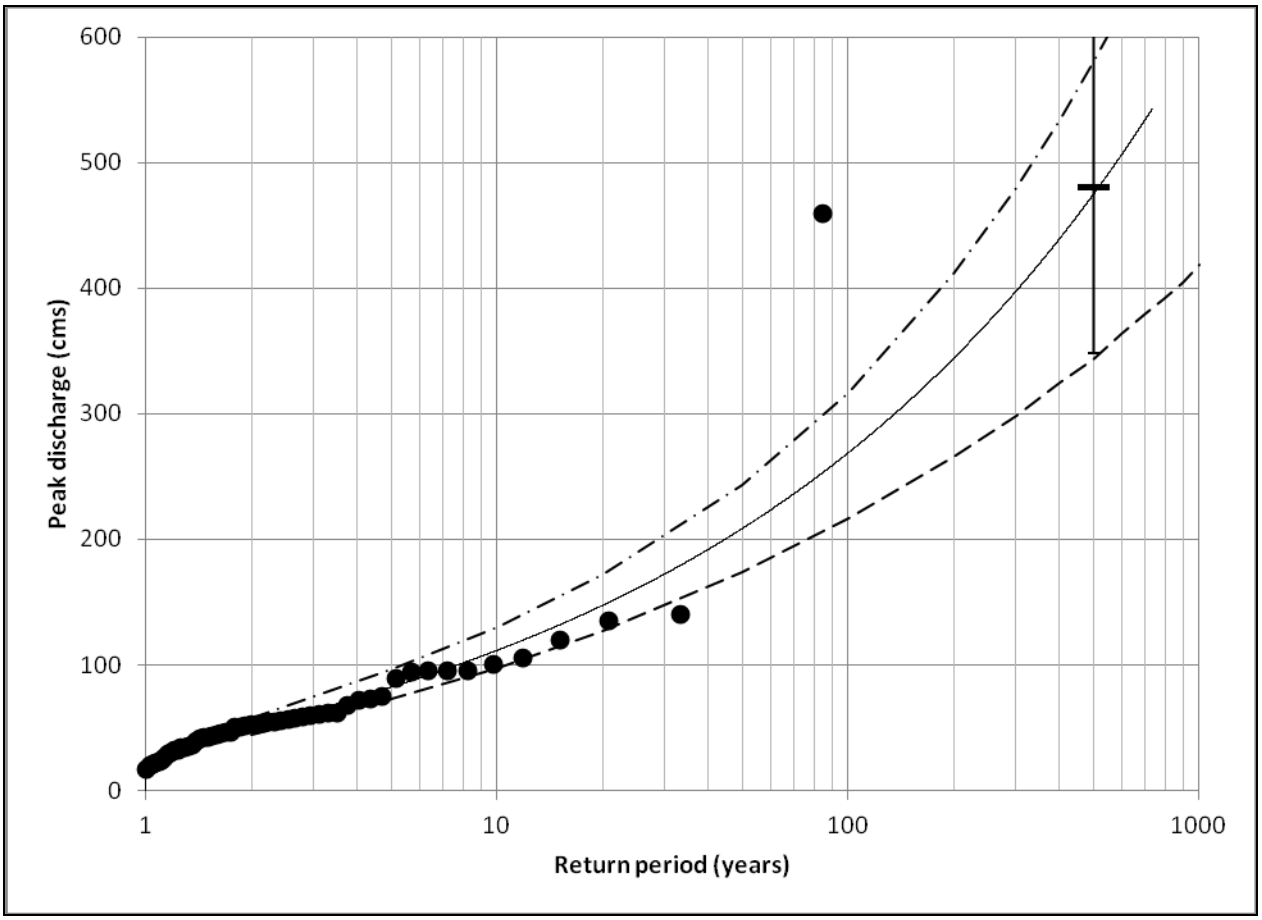

Figure 6. Bayesian fit of the GEV distribution to the systematic data record of the Kamp at Zwettl for 1951–2005, including causal information expansion. The expert guess for Q_{500} is shown along with its 5% and 95% quantiles ($\sigma_{500} = 80\,\mathrm{m^3/s}$). The continuous line corresponds to the PM estimate. The 5% and 95% credible bounds are shown as dashed lines. (cms = m^3/s)

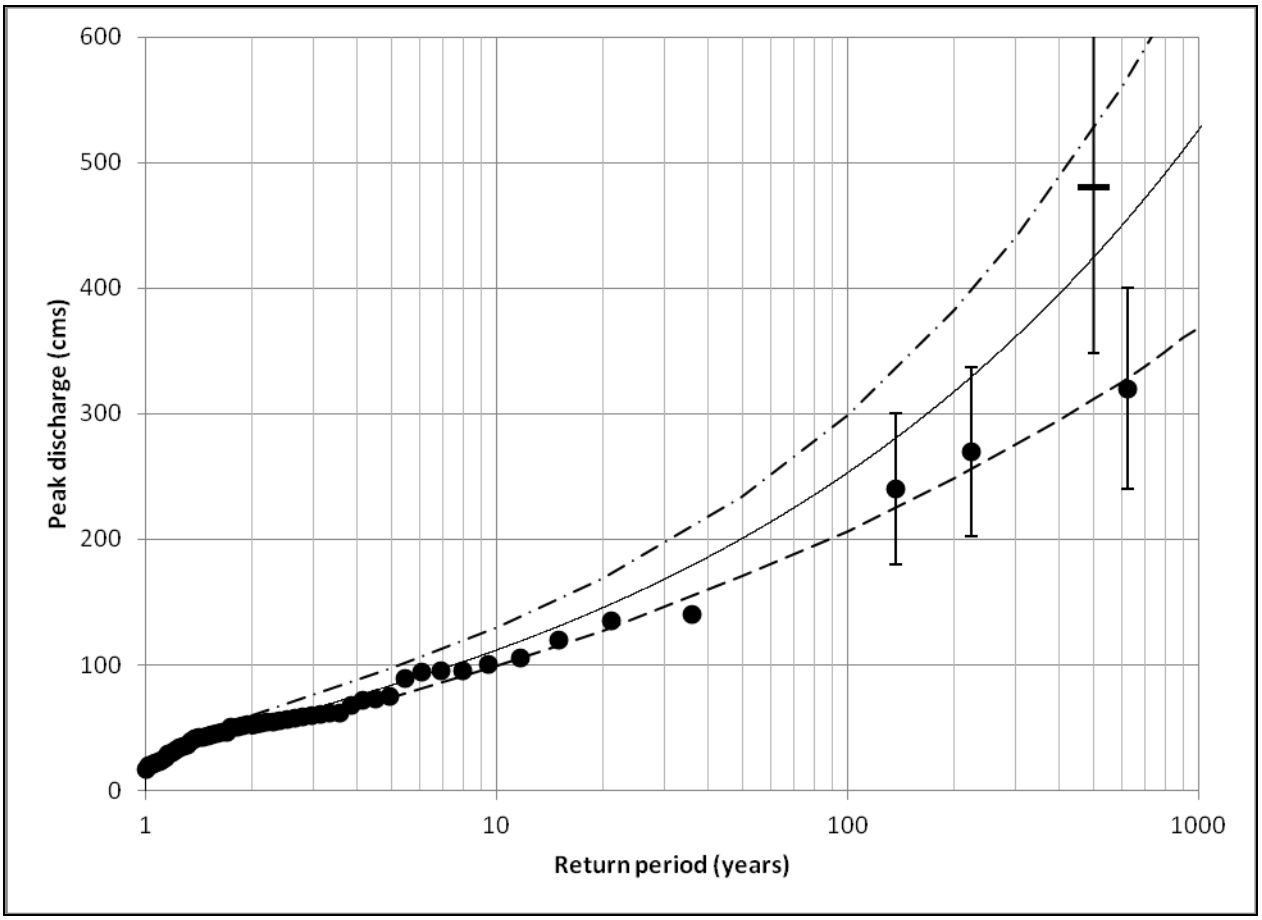

Figure 7. Bayesian fit of the GEV distribution to the systematic data record of the Kamp at Zwettl for 1951–2001, including temporal and causal information expansion data. The temporal information expansion data is shown, including uncertainty. The expert guess for Q_{500} is shown along with its 5% and 95% quantiles ($\sigma_{500} = 80\,\mathrm{m^3/s}$). The continuous line corresponds to the PM estimate. The 5% and 95% credible bounds are shown as dashed lines. (cms = m^3/s)

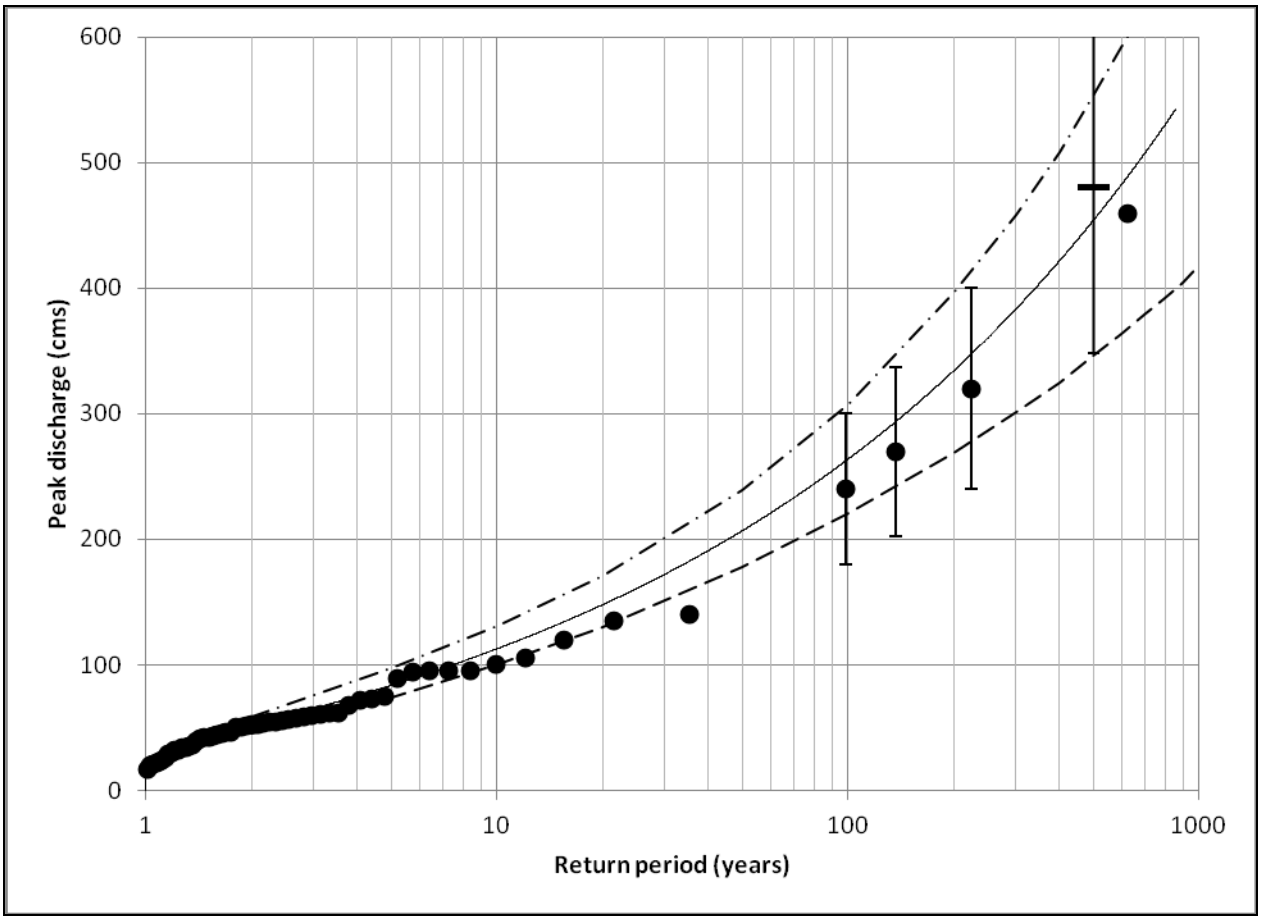

Figure 8. Bayesian fit of the GEV distribution to the systematic data record of the Kamp at Zwettl for 1951–2005, including temporal and causal information expansion data. The temporal information expansion data is shown, including uncertainty. The expert guess for Q_{500} is shown along with its 5% and 95% quantiles ($\sigma_{500} = 80\,\mathrm{m^3/s}$). The continuous line corresponds to the PM estimate. The 5% and 95% credible bounds are shown as dashed lines. (cms = m^3/s)

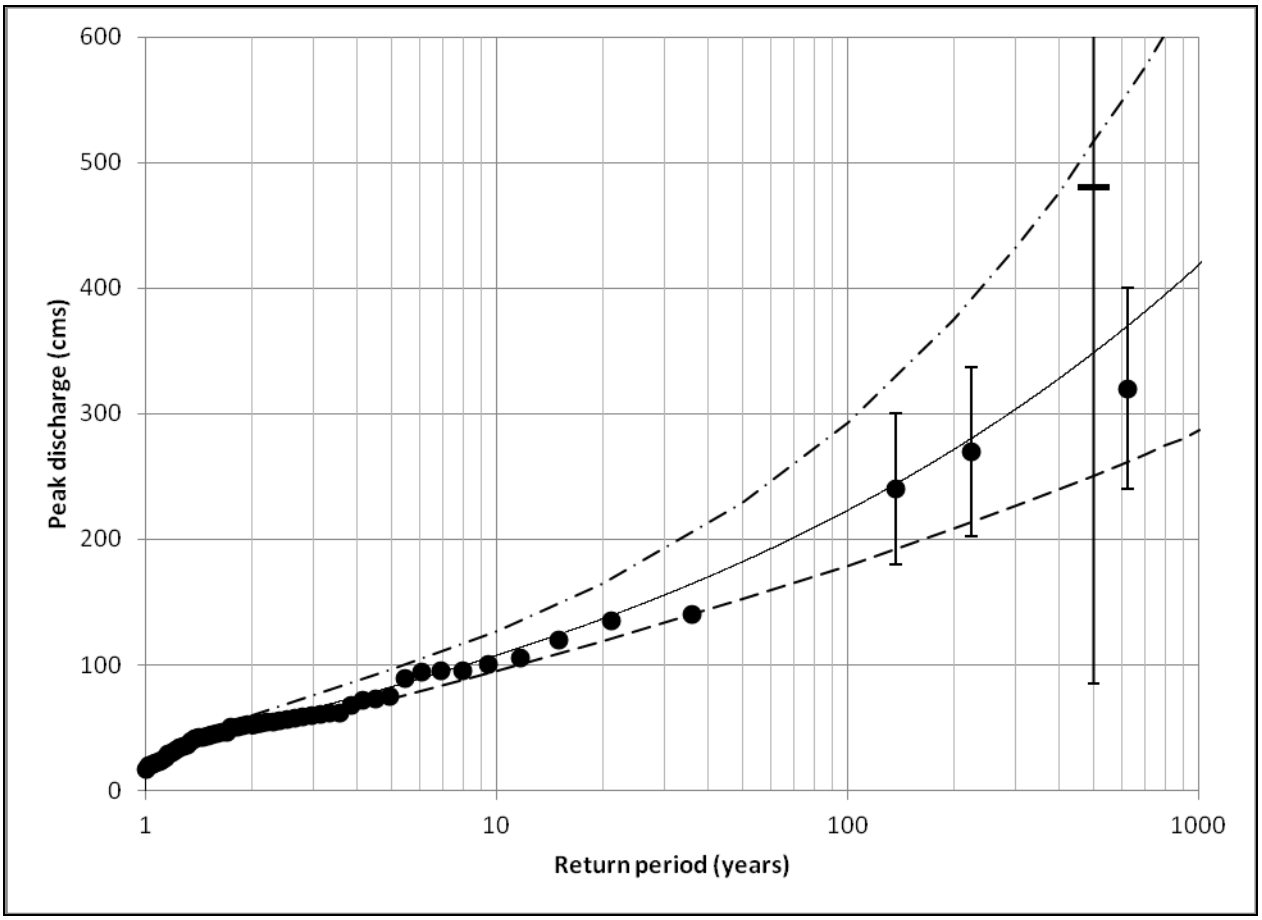

Figure 9. Bayesian fit of the GEV distribution to the systematic data record of the Kamp at Zwettl for 1951–2001, including temporal and causal information expansion data. The temporal information expansion data is shown, including uncertainty. The expert guess for Q_{500} is shown along with its 5% and 95% quantiles ($\sigma_{500} = 80\,\mathrm{m^3/s}$). The continuous line corresponds to the PM estimate. The 5% and 95% credible bounds are shown as dashed lines. (cms = m^3/s)

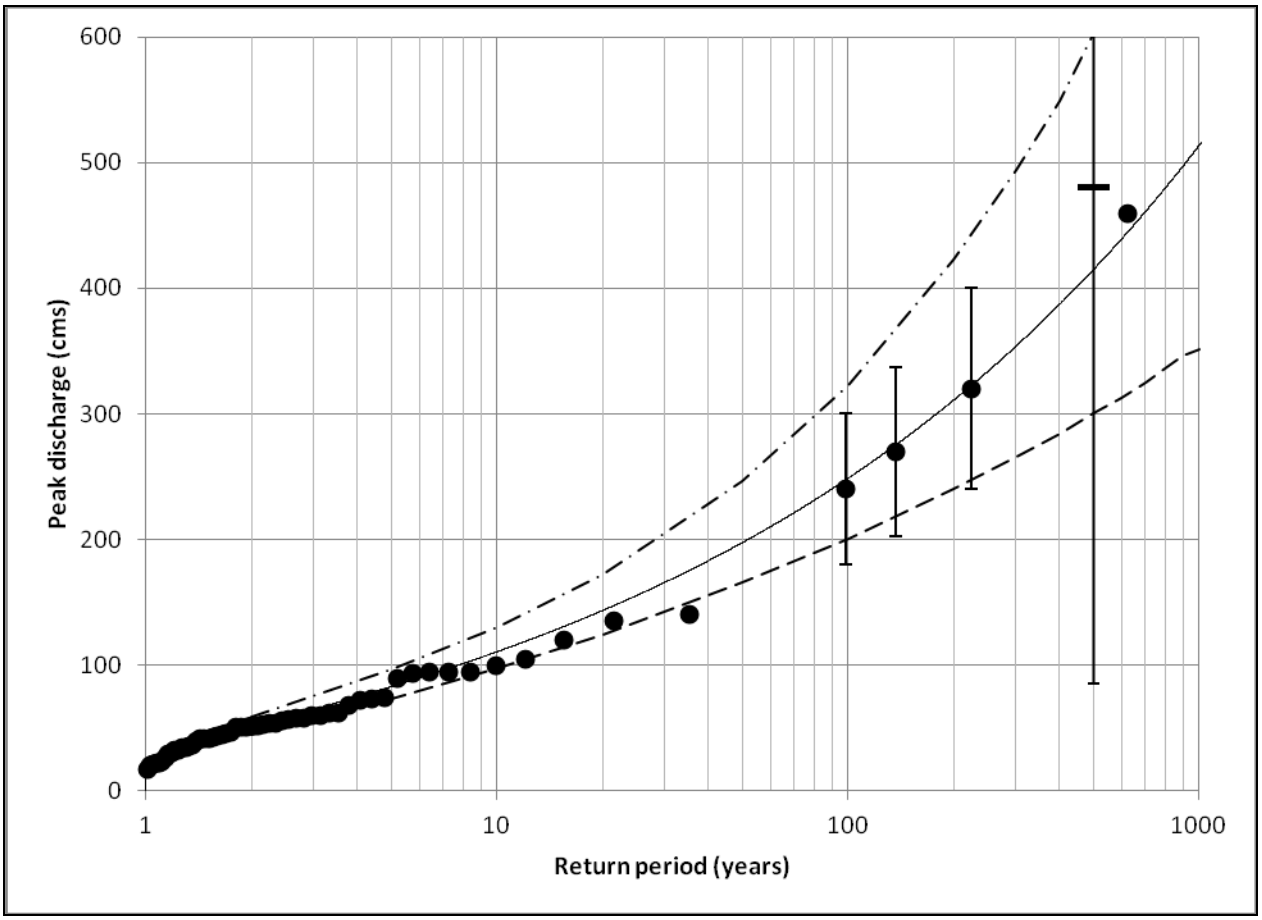

Figure 10. Bayesian fit of the GEV distribution to the systematic data record of the Kamp at Zwettl for 1951–2005, including temporal and causal information expansion data. The temporal information expansion data is shown, including uncertainty. The expert guess for Q_{500} is shown along with its 5% and 95% quantiles ($\sigma_{500} = 240\,\mathrm{m^3/s}$). The continuous line corresponds to the PM estimate. The 5% and 95% credible bounds are shown as dashed lines. (cms = m^3/s)

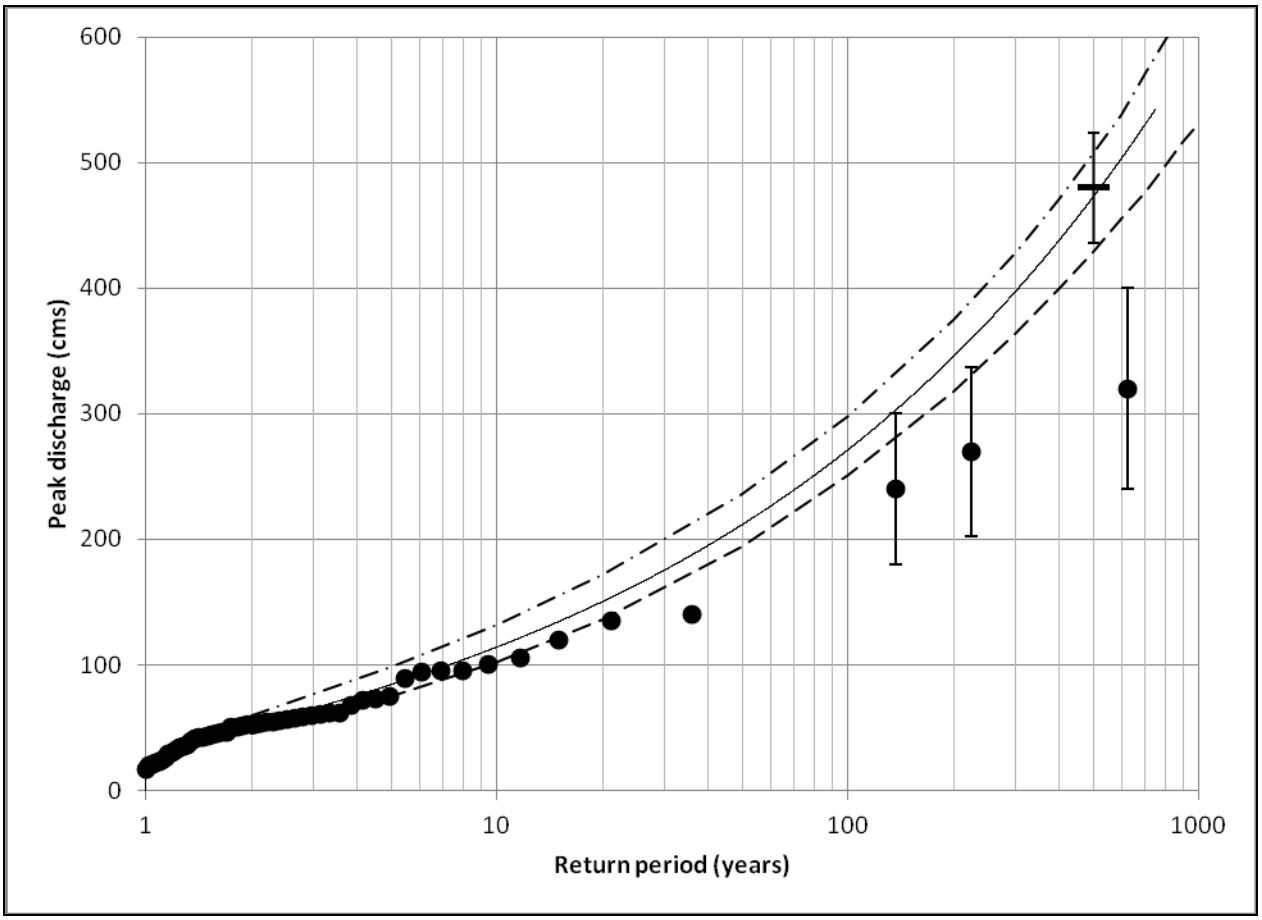

Figure 11. Bayesian fit of the GEV distribution to the systematic data record of the Kamp at Zwettl for 1951–2001, including temporal and causal information expansion data. The temporal information expansion data is shown, including uncertainty. The expert guess for Q_{500} is shown along with its 5% and 95% quantiles ($\sigma_{500} = 26.7$ m³/s). The continuous line corresponds to the PM estimate. The 5% and 95% credible bounds are shown as dashed lines. (cms = m³/s)

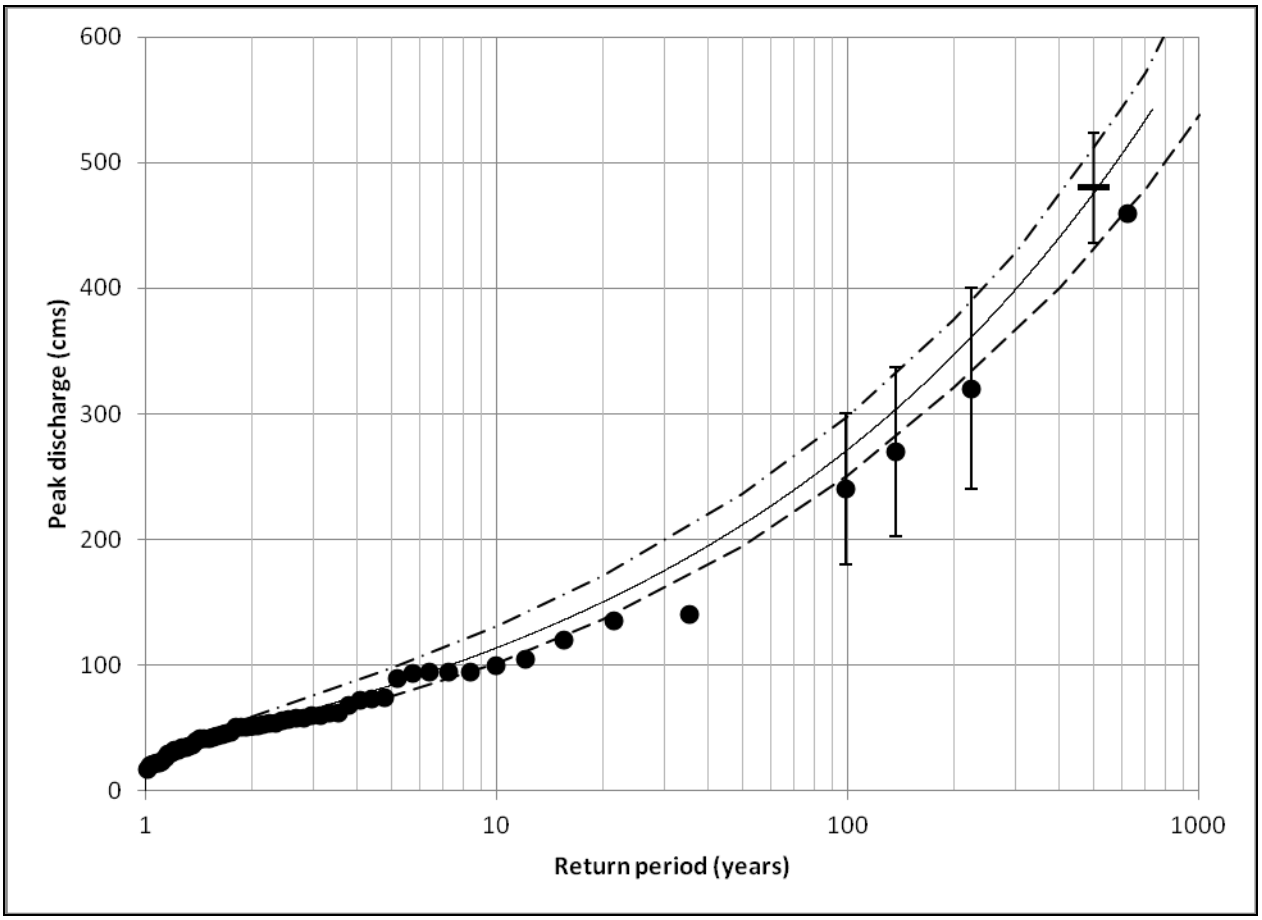

Figure 12. Bayesian fit of the GEV distribution to the systematic data record of the Kamp at Zwettl for 1951–2005, including temporal and causal information expansion data. The temporal information expansion data is shown, including uncertainty. The expert guess for Q_{500} is shown along with its 5% and 95% quantiles ($\sigma_{500} = 26.7$ m³/s). The continuous line corresponds to the PM estimate. The 5% and 95% credible bounds are shown as dashed lines. (cms = m³/s)

DISCUSSION: This technical note has succinctly revisited parts of a Bayesian analysis of the flood frequency hydrology concept originally performed by Viglione et al. (2013) for the 622 km² Kamp at Zwettl river basin located in northern Austria. Eight primary MCMC simulations were performed to examine the impacts of combining different but complementary data sources relevant to flood frequency curve estimation before and after the 2002 flood event in the Kamp, viz., the systematic data record, historic flood information, and expert elicitation for Q_{500} derived from rainfall-runoff modeling analysis in the Kamp, and regionally.

The first two MCMC simulations only considered the systematic data record, until 2001, and also until 2005. Flood quantile estimates, including their computed 90% credible intervals, as depicted in Figures 1 and 2 and also listed in Table 2, differ significantly across these two simulations not only by virtue of the brief record of observations in either case but also because the value of the 2002 flood departs significantly from the remainder of the record. The range of the computed 90% credible intervals plotted in Figures 1 and 2 and also listed in Table 2 for Q_{100} and Q_{1000} clearly underscores a high degree of uncertainty for the flood quantile estimates

associated with each simulation, again, attributed to the short systematic data record. The PM estimates derived from MCMC simulations 1 and 2 for the return period of a flood equal in magnitude to the 2002 flood event is ~85,000 years and 598 years, respectively. These results, albeit derived by only considering the systematic data record, clearly underscore that estimates may quickly become dated, and moreover, the importance of combining additional sources of data relevant to frequency curve estimation.

Including the historical flood information significantly improved upon the agreement of the estimates computed before and after the 2002 flood, including their computed 90% credible interval bounds, more so for return periods less than 100 years than for larger return periods. The PM estimates for Q_{100} differ by 12%; whereas, the PM estimates for Q_{1000} differ by 25%. For MCMC simulations 3 and 4, not only is there better agreement of the their computed 90% credible interval bounds relative to the first two simulations, but the ranges are decreased in each case as well, indicating improved estimation of the GEV distribution parameters by virtue of inclusion of the additional temporal information expansion data into the Bayesian MCMC supervised optimization and inference process. While the PM-based return period estimates from MCMC simulations 1 and 2 for a flood equal in magnitude to the 2002 flood differed by two orders of magnitude, upon consideration of the temporal information expansion data, they now differ by approximately a factor of 2.

Comparison of the results obtained by incorporating the causal information expansion data (i.e., MCMC simulations 5 and 6) with the results from the previous two MCMC simulations that included the historical flood information (i.e., MCMC simulations 3 and 4) indicates improved agreement of the flood quantile estimates before and after the 2002 flood event, including for the larger return periods. In this case (i.e., MCMC simulations 5 and 6), the PM estimates for Q_{100} and Q_{1000} before and after the 2002 flood event differ by 4% and 8%, respectively. The PM-estimated return period values for a flood equal in magnitude to the 2002 flood event now only differ by 23%.

Combining the systematic data record together with both the temporal information expansion and causal information expansion data (i.e., MCMC simulations 7 and 8) did not result in any further improvement with respect to agreement of the flood quantile estimates, including their computed and reported 90% credible interval bounds, before and after 2002, when compared with the results obtained by simply considering the causal information expansion data (i.e., MCMC simulations 5 and 6). However, by also considering the historical flood information together with the systematic data record and the expert's estimate for Q_{500} in the Bayesian MCMC analysis, the influence of the temporal information expansion data is clearly evident upon comparing the results encapsulated in Figure 5/6 with Figure 7/8. In particular, it uniformly shifts the flood quantile estimates slightly toward the historical flood information imparted to the analysis. The flood frequency estimates obtained by combining the three data sources via Bayesian MCMC analysis differ only modestly before and after 2002. In this case, the PM estimated 100-year flood peak at Zwettl before and after 2002 is 253 m^3/s and 264 m^3/s, respectively. The PM-based estimated return periods before and after 2002 for a flood equal in magnitude to the 2002 flood event are 643 years and 517 years, respectively.

Figure 13 and Figure 14 are distributions for Q_{100} and Q_{1000} derived from the Bayesian MCMC simulations 7 and 8. Uncertainty for these two flood quantiles, resultant from the specified component parts of the Bayesian analysis (e.g., the prior distribution and likelihood function), is explicitly presented in these Figures. MCMC simulations 7 and 8 differ only by virtue of the additional 4 years (2002–2005) of the systematic data record included in simulation 8. The observed slight translation to the right for the distributions associated with simulation 8 in Figure 13 and Figure 14 is attributed to the inclusion of the 2002 flood event in the eighth simulation. Of importance, although as previously mentioned dependent upon the component parts of the Bayesian analysis, the statistical difference in estimating these two flood quantiles before and after the 2002 flood event is encapsulated in these two Figures. For example, combining the three different data sources via the Bayesian MCMC simulations 7 and 8, the resultant estimated probability that the 1000-year flood peak at Zwettl is between 450 and 650 m^3/s is 0.68 before 2002 and 0.75 afterwards. The results presented in Figures 13 and 14, associated with MCMC simulations 7 and 8 that combined the three different data sources, provide an opportunity to make two important points regarding the Bayesian analysis approach. These include (1) the flexibility and ease with which it can be employed to coherently combine different types of data, including their uncertainty, relevant to the frequency curve estimation and uncertainty analysis (see Viglione et al. [2013] and references cited therein for additional discussion regarding ways with which to combine different data sources), and (2) that the outcome of its application is a set of random draws from $p(\mathbf{p}|D)$, which can be used to make formal probabilistic-based inferences regarding functions of $\mathbf{p}$, such as the flood quantiles. This approach avoids any reliance on arbitrary computations to derive the final flood quantile estimates from multiple data sources, including their uncertainties, which can potentially confound their meaning. A Bayesian analysis of the flood frequency hydrology concept (Viglione et al. 2013) meshes well with the previously mentioned requirements outlined in existing USACE policy guidance for flood damage reduction studies.

The four additional MCMC simulations (i.e., 9–12), which also combined all three data sources, were designed, simply for purposes of demonstration, to simulate the effect of including *poor* and *excellent* local flood production process understanding relative to the base case originally profiled by Viglione et al. (2013), which was arbitrarily deemed categorically as *good*. MCMC simulations 9/10 and 11/12 as designed differ from simulations 7/8 only by specification of σ_{500}, the specified standard deviation associated with the assumed normal distribution for the 500-year flood runoff, whose mean value is equal to 480 m^3/s. The results associated with MCMC simulations 7/8, 9/10, and 11/12 notably differ. The results associated with simulations 9 and 10, which incorporated the *poor* causal information expansion data, are dominated by the historical flood information for larger return periods; whereas, the results for simulations 11 and 12, which incorporated the *excellent* local flood production process understanding into the Bayesian analysis are dominated by the causal information expansion data at the larger return periods. While simply illustrative, the results obtained from the final four MCMC simulations emphasize the importance to correctly quantify, insofar as is possible, the uncertainties of the data to be combined via the Bayesian analysis.

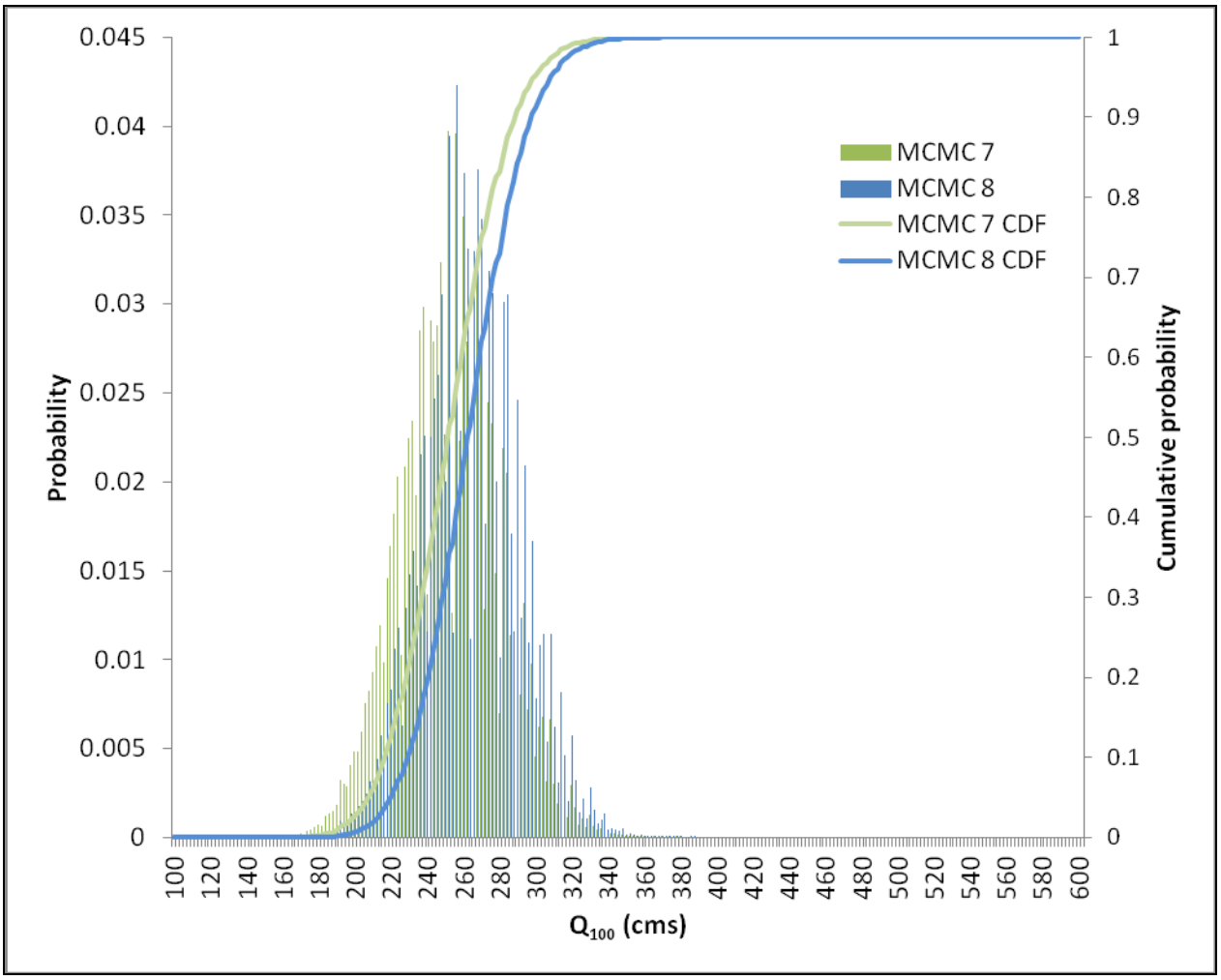

Figure 13. Distributions of Q_{100} associated with MCMC simulations 7 and 8.

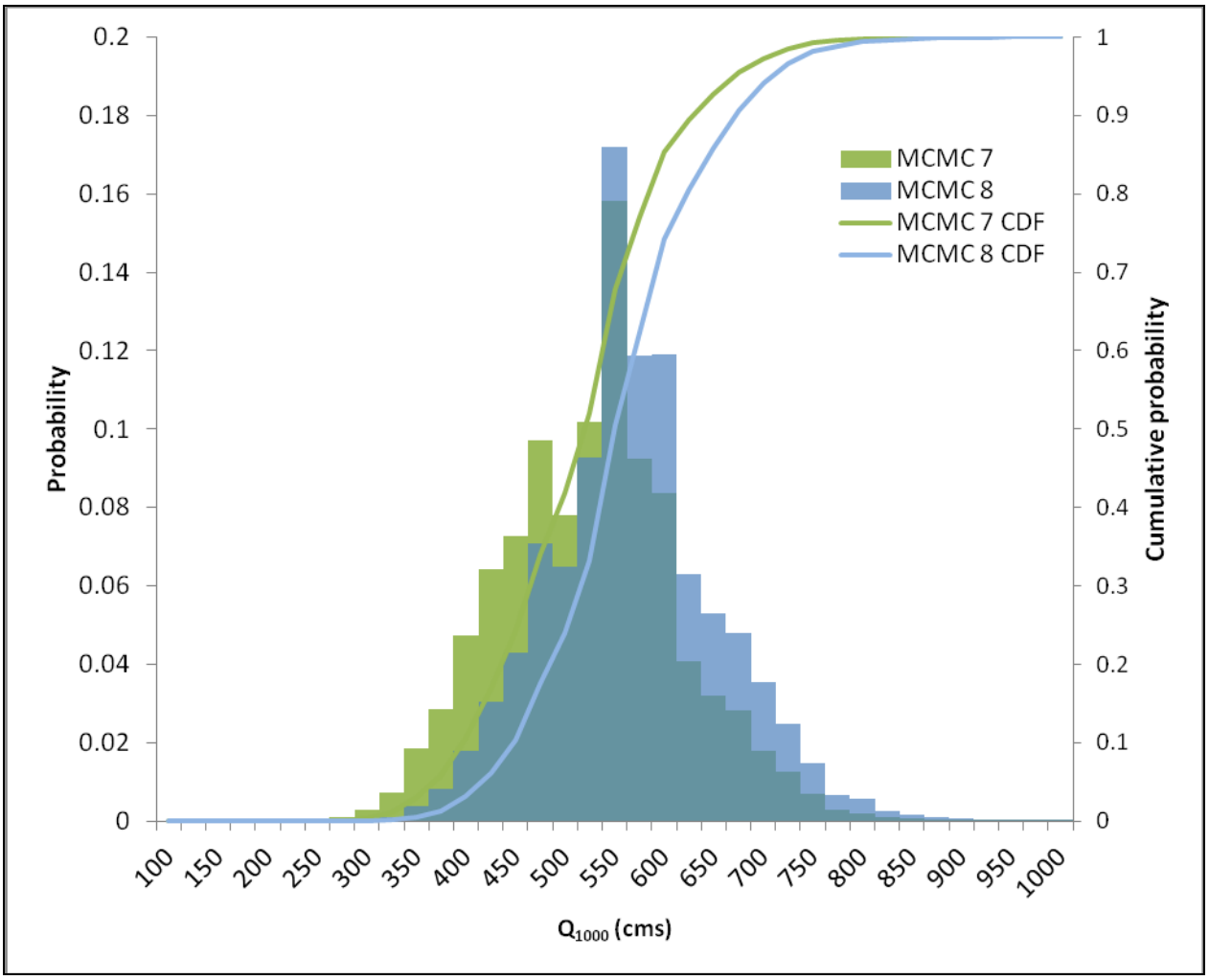

Figure 14. Distributions of Q_{1000} associated with MCMC simulations 7 and 8.

CONCLUSIONS: The content of this document has demonstrated a new USACE capacity to perform a Bayesian analysis of the flood frequency hydrology concept by independently revisiting parts of the example originally profiled by Viglione et al. (2013) for the 622 km^2 Kamp at Zwettl river basin located in northern Austria. A Bayesian analysis of the flood frequency hydrology concept is attractive in that it permits one to flexibly and coherently combine multiple, independent data sources relevant to a flood frequency analysis, including the systematic record, and also, dependent upon availability, temporal, spatial, and causal information expansion data. In addition, its assumptions are made explicit, and the analysis is repeatable and revisable. Moreover, its application provides a basis to make formal probabilistic-based inferences regarding the flood quantiles. A Bayesian analysis of the flood frequency hydrology concept satisfies the requirements of existing USACE policy guidance regarding flood damage reductions studies, viz., a probabilistic analysis of "all key variables, parameters, and components of flood damage reduction studies" (USACE 2006).

In this technical note, multiple Markov Chain Monte Carlo simulations were performed to combine a brief systematic record with historical flood data and information pertaining to local

flood production process understanding, obtained via expert elicitation, in different ways, to simultaneously optimize and infer the posterior distribution of the GEV distribution parameters implied by each Bayesian modeling analysis. It is underscored to the reader for clarity that distributions other than the GEV could easily be considered within the Bayesian analysis framework. Moreover, it is flexible in that the different data sources can be combined in many ways other than the one approach profiled herein (Viglione et al. 2013). While possibly mediated by the consideration of an additional data source, viz., spatial expansion information, the four supplemental MCMC simulations that explored the impact of varying the standard deviation associated with the causal information expansion data source nonetheless underscored the importance of correctly quantifying and assigning the appropriate uncertainty values to the separate pieces of information imparted to the Bayesian analysis.

Two potential related USACE civil works research and development opportunities include the following:

1. Explore the consideration of land use and/or climate change within the Bayesian analysis of the flood frequency hydrology concept framework, likely via causal information expansion data obtained from rainfall-runoff modeling and/or a time dependency treatment of the distribution's parameters.

2. Incorporate an implementation of the Bayesian analysis of the flood frequency hydrology concept into the Hydrologic Engineering Center's Statistical Software Package (HEC-SSP) tool (USACE 2010) (e.g., to support analyses in item 1 directly above and/or a user-defined method option for computing uncertainty when combining multiple data sources).

ADDITIONAL INFORMATION: This CHETN was prepared as part of the Extreme Hydrologic Events work unit in the Infrastructure R&D Program and was written by Dr. Brian E. Skahill (*Brian.E.Skahill@usace.army.mil*) and Dr. Aaron Byrd (*Aaron.R.Byrd@usace.army.mil*) of the U.S. Army Engineer Research and Development Center (ERDC), Coastal and Hydraulics Laboratory (CHL), and Dr. Alberto Viglione (*viglione@hydro.tuwien.ac.at*) of the Institute of Hydraulic Engineering and Water Resources Management at the Vienna University of Technology. The Program Manager is Dr. Cary Talbot, and the Technical Director is William Curtis. This CHETN should be cited as follows:

> Skahill, B. E., A.Viglione, and A. R. Byrd. 2016. *A Bayesian analysis of the flood frequency hydrology concept*. ERDC/CHL CHETN-X-1. Vicksburg, MS: U.S. Army Engineer Research and Development Center. *http://chl.erdc.usace.army.mil/chetn*

REFERENCES

Coles, S. G., and J. A. Tawn. 1996. A Bayesian analysis of extreme rainfall data. *J. R. Stat. Soc., Ser. C Appl. Stat.* 45(4):463–478.

Dalrymple, T. 1960. *Flood frequency analysis.* U.S. Geol. Surv. Water Supply Pap. 1543A.

Duband, D., C. Michel, H. Garros, and J. Astier. 1994. *Design flood determination by the Gradex method*. CIGB, Int. Comm. On Large Dams, Paris.

Eagleson, P. S. 1972. Dynamics of flood frequency. *Water Resour. Res.* 8(4):878–898. doi:10.1029/WR008i004p00878

Gelman, A., and D. B. Rubin. 1992. Inference from iterative simulation using multiple sequences. *Stat. Sci.* 7:457–472.

Guillot, P. 1972. Application of the method of Gradex, in floods and droughts. In *Proceedings of the Second International Symposium in Hydrology.* Ed. E. F. Schulz, V. A. Koelzer, and K. Mahmood, 44–49. Fort Collins, CO: Water Resource Publications.

Hastings, W. K. 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57:97–109.

Merz, R., and G. Blöschl. 2008a. Flood frequency hydrology: 1. Temporal, spatial, and causal expansion of information. *Water Resour. Res.* 44:W08432. doi:10.1029/2007WR006744

Merz, R., and G. Blöschl. 2008b. Flood frequency hydrology: 2. Combining data evidence. *Water Resour. Res.* 44:W08433. doi:10.1029/2007WR006745

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. 1953. Equation of state calculations by fast computing machines. *J. Chem. Phys.* 21:1087–1092.

Naghettini, M., K. W. Potter, and T. Illangasekare. 1996. Estimating the upper tail of flood-peak frequency distributions using hydrometeorological information. *Water Resour. Res.* 32(6):1729–1740. doi:10.1029/96WR00200

Rahman, A., P. E. Weinmann, T. M. T. Hoang, and E. M. Laurenson. 2002. Monte Carlo simulation of flood frequency curves from rainfall. *J. Hydrol.* 256:196– 210. doi:10.1016/S0022-1694(01)00533-9

Sivapalan, M., E. F. Wood, and K. Beven. 1990. On hydrologic similarity: 3. A dimensionless flood frequency model using a generalized geomorphologic unit hydrograph and partial area runoff generation. *Water Resour. Res*. 26(1):43– 58.

Sivapalan, M., G. Blöschl, R. Merz, and D. Gutknecht. 2005. Linking flood frequency to long-term water balance: Incorporating effects of seasonality. *Water Resour. Res.* 41:W06012. doi:10.1029/2004WR003439

ter Braak, C. J. F., and Vrugt, J. A. 2008. Differential evolution Markov chain with snooker updater and fewer chains. *Stat. Comput.* 18(4):435 – 446. doi: 10.1007/s11222-008-9104-9

U.S. Army Corps of Engineers (USACE). 1996. *Risk-based analysis of flood damage reduction studies*. EM 1110-2-1619. Washington, DC.

_____. 2006. *Risk based analysis for evaluation of hydrology/hydraulics, geotechnical stability, and economics in flood damage reduction studies.* ER 1105-2-101. Washington, DC.

_____. 2010. *HEC-SSP statistical software package user's manual, version 2.0*. Institute for Water Resources. Hydrologic Engineering Center. Davis, CA.

Viglione, A., R. Merz, J. L. Salinas, and G. Blöschl. 2013. Flood frequency hydrology: 3. A Bayesian analysis, *Water Resour. Res.* 49. doi:10.1029/2011WR010782

Wiesbauer, H. 2004. *Historische überflutungen am Kamp*. Technical report. Osterreich, Germany: Wasser Niederösterreich, St. Pölten.

Wiesbauer, H. 2007. *Historische überflutungen am Kamp, in extreme abflussereignisse, Wiener Mitt.* Vol. 206. Ed. D. Gutknecht, 43–58. Vienna, Austria: Inst. für Wasserbau und Ingenieurhydrol.